Article:

Bolibaugh, Cylcia orcid.org/0000-0001-7500-264X and Foster, Pauline (2021) Implicit Statistical Learning in Naturalistic and Instructed Morphosyntactic Attainment: An Aptitude-Treatment Interaction Design. Language Learning. pp. 959-1003. ISSN 0023-8333

https://doi.org/10.1111/lang.12465

LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Implicit Statistical Learning in Naturalistic and Instructed Morphosyntactic Attainment: An Aptitude-Treatment Interaction Design

Cylcia Bolibaugh [a] and Pauline Foster [b]

[a]University of York [b]St. Mary's University, Twickenham, London

Abstract: We investigated the potential influence of implicit learning mechanisms on L2 morphosyntactic attainment by examining the relationship between age of onset (AoO), two cognitive abilities hypothesized to underlie implicit learning (phonological short-term memory and implicit statistical learning), and performance on an auditory grammaticality judgment test (GJT). Participants were 71 Polish-English long-term bilinguals with a wide range of AoOs (1–35 years) who differed in their context of learning and use (immersed vs. instructed). In immersed learners, we observed a growing dissociation between performance on grammatical and ungrammatical sentences as AoO was delayed. This effect was attenuated in those with better phonological short-term memory and statistical learning abilities and is consistent with a decline in the ability to learn from implicit negative evidence. In instructed learners, GJT performance was subject to additive effects of AoO and grammaticality and was not associated with either cognitive predictor, suggesting that implicit learning mechanisms were not involved.

Keywords age effects; statistical learning; implicit learning; phonological short-term memory; grammaticality judgment

This study was funded by a grant from the U.K. Economic and Social Research Council, reference RES-000-22-2645.

Correspondence concerning this article should be addressed to Cylcia Bolibaugh, Centre for Research in Language Learning and Use, Department of Education, University of York, YO10 5DD, UK. E-mail: cylcia.bolibaugh@york.ac.uk

The handling editor for this article was Kara Morgan-Short.

DOI: 10.1111/lang.12465

Introduction

Understanding why age of onset (AoO) is negatively correlated with language attainment represents one of the most substantive empirical problems of second language acquisition (SLA). One now well-established interpretation is that age effects reflect the loss of implicit learning ability (DeKeyser, 2000; Paradis, 2009), with the strongest form of the proposal entailing a qualitative shift in learning mechanisms in childhood: "[S]omewhere between the ages of 6–7 and 16–17, everybody loses the mental equipment required for the implicit induction of the abstract patterns underlying a human language" (DeKeyser, 2000, p. 518). The claim of a fundamental difference between child and adult language learning (Bley-Vroman, 1990) has drawn on evidence from the interactions of AoO with learning context and with aptitude. Age effects have most reliably been found in immersed contexts where input conditions favor implicit learning, whereas individual differences in verbal analytic ability, or explicit aptitude, have been found to mediate attainment in late but not early-onset naturalistic learners (DeKeyser, 2000; DeKeyser, Alfi-Shabtay, & Ravid, 2010).

Though the shift-of-learning-mechanisms explanation for age effects thus reconciles findings from several seminal SLA studies, it has not been consistently corroborated. Counter to the prediction that adults should no longer be able to implicitly induct patterns, experimental research using (semi-) artificial language systems has reported evidence of adults' learning regularities in the absence of awareness (e.g., Grey, Williams, & Rebuschat, 2014; Marsden, Williams, & Liu, 2013; Rebuschat & Williams, 2012; Williams & Kuribara, 2008). Other studies of long-term attainment that have reported benefits for an early start in instructed contexts (Larson-Hall, 2008) and benefits for explicit aptitude in early-onset learners (Abrahamsson & Hyltenstam, 2008; Granena & Long, 2013) further weaken the premise that age should only matter in immersion contexts and the notion that explicit aptitude is *only* necessary to compensate for a loss of implicit learning ability in older learners. Finally, perhaps the most striking corollary of a strong version of the fundamental difference proposal, namely, that individual differences in implicit learning ability should *not* mediate ultimate attainment for those who start learning in adulthood, has only been indirectly examined in a single study of ultimate attainment, that of Granena (2013). Granena's report of associations between early- and late-onset second language (L2) learners' morphosyntactic attainment and probabilistic sequence learning—an inductive statistical learning mechanism proposed to underlie native language acquisition (Erickson & Thiessen, 2015; Jimenez, 2003)—raises further challenges for the fundamental difference account.

The aim of the present study was to address these inconsistencies and challenges by extending the research that has assessed how far and in what conditions statistical learning mechanisms influence morphosyntactic attainment in immersed and instructed learners. We did so by adopting a similar approach to that of previous studies of ultimate attainment, using an aptitude-treatment interaction design (DeKeyser, 2012). Specifically, we explored the ultimate morphosyntactic attainment, as assessed by a grammaticality judgment test (GJT), of long-term L2 users in relation to AoO and two cognitive abilities hypothesized to underlie implicit learning: phonological short-term memory (PSTM) and implicit statistical learning (ISL). Unlike what previous studies did, we assessed these relationships in both immersed and instructed learners who were long-term, daily users of their L2. Finally, by measuring individual difference variables that underlie systematic differences in the development of implicit knowledge, we also addressed recent debates regarding the construct validity of GJTs and the role of item grammaticality.

Background Literature

A recurring issue in the field of L2 learning is the difficulty in measuring implicit knowledge (Erlam, 2006). We therefore begin by reviewing criteria that have been used to diagnose implicit knowledge and consider evidence for implicit learning from laboratory studies that have used similar tests to those used in studies of ultimate attainment before we review the growing validation literature to address the possibility of measuring implicit knowledge via GJTs. We next contextualize our design by examining findings from two interactions (AoO with learning context and AoO with aptitude) that have been used to make claims about the scope and provenance of age effects in L2 learning, and we assess the extent to which they support claims of qualitative differences between child and adult L2 learning. Finally, in order to motivate our choice of individual difference measures, we set out a memory-based account of ISL.

Implicit Knowledge and Learning of Morphosyntax in Second Language Acquisition

The criteria most commonly used to operationalize implicit knowledge in SLA were synthesized by R. Ellis for use in the large-scale Marsden Project (R. Ellis, 2005). Researchers have agreed that implicit language knowledge is tacit and intuitive (Reber, 1989). As the cornerstone of linguistic competence, it enables even young children to understand what is possible in their language before they develop metalinguistic insight into why that might be so (e.g., see Karmiloff-Smith, 1979, on epilinguistic awareness). Implicit knowledge is also procedural,[1] and so it influences behavior in the absence of awareness (Cleeremans, Destrebecqz, & Boyer, 1998). Finally, a feature that is perhaps less evident is that implicit knowledge is more systematic, or structured, than explicit knowledge (R. Ellis, 2004; Tarone, 1988). It is a combination of systematicity and automaticity that underlies two key behavioral hallmarks of implicit language knowledge, namely, that "[p]eople have fluent and productive command of their native language and are able to instantly detect grammatical irregularities, without being able to explain the underlying rules" (Williams, 2009, p. 319).

The value of implicit linguistic *knowledge* may be clear, but what evidence is there for implicit second language *learning*?[2] One productive strand of research has made use of semi-artificial language paradigms in which exposure conditions and input can be strictly controlled. Implicit learning in these studies has most often been operationalized as above-chance performance on judgment tasks, in the absence of awareness of what has been learned, following incidental exposure to language-like regularities. A series of such studies simulating the learning of L2 morphosyntax under cover of meaning-focused tasks indicated that learning does occur under such incidental conditions and that at least part of the knowledge acquired is unconscious and abstract, with a phenomenology similar to native-speaker intuition (Grey et al., 2014; Marsden et al., 2013; Rebuschat & Williams, 2012; Williams & Kuribara, 2008). These results suggested a retained capacity for implicit learning, but they also pointed to possible limits of learning from mere exposure in adults: Although participants consistently demonstrated learning for both previously encountered and novel grammatical items, they failed to reject ungrammatical items at levels above chance. Rebuschat and Williams (2013) themselves concluded that what is learned is therefore probably more akin to a series of syntactic patterns rather than categorical linguistic rules. Results such as these (see also, e.g., Andringa & Curcic, 2015, among others) have prompted R. Ellis (2005, 2009) to argue that rejection of ungrammatical items in L2 judgment tasks may require the use of explicit knowledge.

Grammaticality Judgment Tests, Grammaticality, and the Measurement of Implicit Knowledge

Attempts to understand whether implicit learning mechanisms are retained in adults are complicated not only by difficulties in measuring awareness but also in understanding how language task features predispose learners to make use of explicit or implicit knowledge. Questions regarding the construct validity of GJTs in particular as measures of implicit or explicit knowledge have a long history. One feature that has been highlighted as likely to influence the use of implicit knowledge is modality: Whereas written GJTs enable backtracking and focus on forms, auditory GJTs require online processing and may promote a focus on meaning (Loewen, 2009). Although most validation studies modeled on the Marsden Project (R. Ellis, 2005) have used written GJTs (cf. Bialystok, 1979, for an early example with auditory stimuli), a small number of studies have found that GJTs with auditory stimuli are both more difficult (Johnson, 1992; Plonsky, Marsden, Crowther, Gass, & Spinner, 2020; Shiu, Yalçın, & Spada, 2018) and more likely to load on factors corresponding to implicit knowledge (Kim & Nam, 2017; Spada, Shiu, & Tomita, 2015) than those with written stimuli. Timed tests in which speeded responses limit the controlled processing required to access explicit knowledge (Godfroid et al., 2015) have also been found to load more highly on factors representing implicit knowledge relative to unspeeded tests (Bowles, 2011; R. Ellis, 2005; Gutiérrez, 2013; Loewen, 2009; Zhang, 2015). Finally, sentence grammaticality has been proposed as playing an important role. Performance on grammatical sentences has also more consistently loaded on implicit knowledge factors (R. Ellis, 2009; Gutiérrez, 2013), whereas ungrammatical sentences have tended to load alongside tests representing explicit knowledge (cf. Kim & Nam, 2017).

A smaller number of recent studies have questioned whether performance on *any* type of GJT can serve as a measure of implicit knowledge. Adopting a similar psychometric design to that of previous studies, Vafaee, Suzuki, and Kachinske (2017) suggested that all the ungrammatical GJT items tapped explicit knowledge, whereas the factor that they named implicit knowledge comprised only performance on a self-paced reading measure and on a word-monitoring task. Suzuki and DeKeyser (2017) reached a similar conclusion and argued that GJTs are too insensitive to serve as a measure of implicit knowledge and instead reflect what they termed automatized explicit knowledge that results from the proceduralization of declarative knowledge, for example, explicit grammatical rules.

Conflicting results such as these may be at least partly attributable to the fact that the focus of validation studies has been almost exclusively on task design features, with little consideration of the types of knowledge representations that learners bring to the task (cf. Philp, 2009). With the exception of Bowles (2011), validation studies have tested learners with primarily classroom learning experience and with relatively little, late, or no immersive experience. Suzuki and DeKeyser's proposal for the primary role of automatized explicit knowledge in GJT performance must similarly be considered in light of the fact that neither native speakers, nor many heritage speakers, nor most immersed immigrant learners start learning with declarative knowledge of grammatical rules. Although the weight of evidence has thus suggested that performance on grammatical items under timed conditions is likely to draw on implicit knowledge, with some evidence that ungrammatical items measured on timed auditory GJTs do so as well (Kim & Nam, 2017), further research systematically considering the learning histories of participants has been needed.

Ultimate Attainment in Second Language Acquisition

Age by Context Interactions

The suggestion that age effects in ultimate attainment can be best understood as a loss of implicit learning ability largely rests on evidence from two types of interactions: AoO with context of learning and AoO with aptitude, though it is likely that the salience of structures presents a third type of interaction (DeKeyser, Alfi-Shabtay, Ravid, & Shi, 2017). It is well established that the processes and outcomes of SLA are conditioned by the context of acquisition because this context influences the type of input, the quantity of input, and the usage patterns of the input, among other things. Studies of ultimate attainment in immigrants have consistently found strong negative correlations between AoO and auditory GJT performance (whole group correlations ranging between $r(22) = -.59$ in McDonald, 2000, and $r(74) = -.80$ in Study 1 from DeKeyser et al., 2010). The findings from studies of age in instructed contexts, though fewer in number, point in the opposite direction. Studies from the Barcelona Age Factor Project (see García Mayo & Lecumberri, 2003; Muñoz, 2011, 2014) exploited a change in curriculum to investigate the influence of starting age of instruction on eventual L2 attainment. No advantage of an early start was found across a range of linguistic outcomes. Similar results have been reported from large-scale investigations of early (AoO = 8 years), and late (AoO = 13 years) learners in Germanophone Switzerland, where late learners surpassed early starters across a range of measures (Pfenninger & Singleton, 2017, 2019), and in cohorts separated by a smaller gap in Germany (Baumert, Fleckenstein, Leucht, Köller, & Möller, 2020; Jaekel, Schurig, Florian, & Ritter, 2017). In contrast to these findings, Larson-Hall (2008), using a GJT similar to those commonly employed in immersion studies, observed a moderate negative association, $r(60) = -.42$, between AoO and GJT performance in Japanese learners of English with substantial exposure. Given the small number of studies and the concentration of existing research in a few seminal projects, questions remain about the length of exposure and the amount of exposure that are sufficient to constitute ultimate attainment in instructed contexts.

Age by Aptitude Interactions

A qualitative change between child and adult learning further implies that high-achieving adult-onset learners must necessarily rely on explicit learning mechanisms. Support for this position comes from findings that ultimate attainment was constrained by explicit aptitude in adult-onset but not child-onset learners. Explicit aptitude has most commonly been operationalized as language analytic ability (Skehan, 1998), or learners' ability to deliberately identify patterns in linguistic input and to infer rules. DeKeyser's (2000) study was the first to explicitly set out to test this hypothesis as an instantiation of Bley-Vroman's fundamental difference hypothesis. Administering an auditory GJT (adapted from Johnson & Newport, 1989), and a Hungarian translation of Part IV of the MLAT (a grammatical inferencing test of language analytic ability) to long-term Hungarian immigrants to the United States varying in AoO, he observed an interaction of AoO and aptitude in GJT performance. Although the AoO and the GJT performances were strongly negatively correlated, all (but one) adult-onset learners who scored in or near native-speaker range had above average aptitude. No such relationship was present in child-onset learners. Although DeKeyser and colleagues replicated these findings in two further studies with long-term immigrants in Israel and the United States (DeKeyser et al., 2010), a more complex picture has emerged from three additional studies employing similar designs. In a sample of L2 users screened for imperceptible nonnativeness, Abrahamsson and Hyltenstam (2008) replicated the relationship between aptitude and GJT performance in adult-onset learners, but also found a strong and significant relationship in child-onset learners, $r(16) = .70$, $p < .001$. In further contrast to both DeKeyser (2000) and Abrahamsson and Hyltenstam (2008), two studies by Granena failed to find a relationship between time-pressured auditory GJT performance and aptitude, both in adult-onset participants (Granena, 2013) and in early-, mid-, and late-onset participants (Granena & Long, 2013). Thus, although early studies of the interaction of explicit aptitude and age strongly supported a qualitative difference between early- and late-onset learners, more recent findings have led researchers to propose a weaker formulation, one of a difference in degree rather than a difference in kind between child and adult learners (Granena & Long, 2013).

Far fewer studies have examined the role of implicit aptitudes and their interaction with AoO in ultimate attainment. In contrast to experimental studies, the ex post facto design of attainment studies has precluded an examination of awareness at the time of learning. As a result, studies of this type have limited the scope of their investigation to finding evidence for data-driven, or statistical, learning mechanisms, which Williams (1999) characterized as learning that proceeds inductively and unconsciously but may result in either implicit or explicit knowledge. The construct of implicit learning aptitude has most recently been operationalized as performance on LLAMA D, a subtest of the LLAMA (Meara, 2005) aptitude battery that measures auditory sequence learning via an explicit recognition test (Granena, 2013), and performance on probabilistic serial reaction time tasks that measure visual pattern learning (Granena, 2013; Kaufman et al., 2010; Suzuki & DeKeyser, 2017).

Two recent studies in naturalistic and instructed learners respectively have investigated the role of implicit learning aptitude in L2 learners' grammatical attainment. Granena (2013) administered both LLAMA D and a probabilistic serial reaction time task alongside measures of metalinguistic knowledge and an online word monitoring test, to early-onset (AoO $<$ 7 years) and late-onset (AoO $>$ 16 years) Chinese first language (L1), long-term immigrants in Spain. In contrast to the predictions of the fundamental difference hypothesis, not only did Granena find that both sequence learning measures mediated long-term grammatical attainment but also that these relationships were present in early and late learner groups, albeit in different dependent measures. Using a similar design, Suzuki and DeKeyser (2017) administered a battery of form and meaning-focused linguistic measures alongside a probabilistic serial reaction time task, a PSTM task, and an explicit aptitude task to adult-onset, Chinese L1 learners of Japanese living in Japan. Performance on the serial reaction time task did not share any variance with either online or form-focused measures of morphosyntactic attainment.

In summary, the results of the single implicit aptitude by age study in naturalistic learners (Granena, 2013) has further tempered strong claims of qualitative differences between early- and late-onset learners, but the results were not directly comparable to studies of explicit aptitude by age due to the use of different dependent measures (Granena did not use a GJT). Suzuki and DeKeyser's (2017) findings suggested that adult-onset instructed learners do not rely on implicit learning mechanisms even when in the target language environment, but the generalizability of these results is limited by the participants' relatively short period of immersive exposure (minimum = 24 months, $M = 47$ months) and their long history of instructed learning.

Implicit Statistical Learning and Language Acquisition

In the present study, we adopted an operationalization of implicit learning derived from the artificial grammar learning paradigm developed by

Reber (1967) and later modified by other researchers (Conway, Bauernschmidt, Huang, & Pisoni, 2010; Conway, Karpicke, & Pisoni, 2007; Karpicke & Pisoni, 2004, among others). Artificial grammar learning tasks traditionally require participants to memorize strings of letters or symbols without the participants' being told these contain hidden regularities generated by an artificial grammar. Participants are then asked to classify new strings according to whether or not the strings follow the grammar. Implicit learning is evidenced when participants classify strings at levels above chance without their being able to describe the rule system that underpins the structures. Although reliably demonstrating implicit learning at the group level, this reflection-based methodology has been criticized (Christiansen, 2019) for recruiting explicit decision-making processes that introduce unsystematic variance that becomes especially problematic in individual difference studies (Siegelman, Bogaerts, Christiansen, & Frost, 2017).

An alternative to requiring participants to make metacognitive judgments about the patterns to which they have been exposed is to measure the improvement in processing that results from implicit learning (Jiménez, Mendez, & Cleeremans, 1996). Applying this approach to Reber's artificial grammar learning paradigm, Karpicke and Pisoni (2004) hypothesized that implicit knowledge of an artificial grammar would improve immediate memory for novel sequences generated by that grammar. They tested the hypothesis using a memory game procedure (modeled after the Milton Bradley electronic game Simon™) that they had previously developed to measure short-term memory in deaf children (Cleary, Pisoni, & Geers, 2001). Under the guise of the memory game, participants were asked to reproduce sequences of colors presented in one of three stimulus modalities: visual only, auditory only, and visual and auditory. Karpicke and Pisoni (2004) not only found a robust learning effect, whereby memory span for grammatically generated sequences was significantly higher than memory for control sequences, but also found a lack of verbalizable knowledge. In posttask verbal reports, very few participants (14/120) reported any degree of awareness of an underlying rule system, and none of the 120 participants, including those who reported awareness, was able to describe the underlying regularities (e.g., "green could follow blue, but red could not follow blue," p. 962). Participants' confidence in their knowledge was similarly very low, with mean confidence ratings on a 5-point scale of 1 (*not confident*) to 5 (*completely confident*) ranging from 1.25 to 1.33 across conditions.

Karpicke and Pisoni (2004) thus concluded that a sequence-reproduction task originally designed to measure short-term memory could be successfully adapted to simultaneously index individual differences in implicit learning.

Although the fundamental logic of using memory tasks to detect learning can be traced to early work by Miller (1956) and Melton (1963), the paradigm termed *statistically induced chunking recall* has been developed in recent work by a group of researchers (Isbilen, Frost, Monaghan, & Christiansen, 2018; Isbilen, McCauley, Kidd, & Christiansen, 2017). Conway et al. (2010) similarly characterized such learning not only as implicit but also statistical. In a further series of experiments linking performance on the memory game task with native language processing ability, Conway et al. concluded that performance on the memory game reflects "sensitivity to the underlying statistical structure contained in sequential patterns" (p. 365). The mechanisms supporting improvement in memory span for grammatical sequences are likely the same as those enabling successful test performance in other forms of artificial grammar learning where "learning is statistical in the sense [of] encoding … the frequency of chunks of elements (Perruchet & Pacteau, 1990) or … learning the transitional probabilities among consecutive elements (Saffran, Johnson, Aslin, & Newport, 1999)" (Conway & Christiansen, 2006, p. 905). Although the nature of the representations learned in the memory game task has not been conclusively established—for example, Karpicke and Pisoni (2004) found that chunk strength did not contribute to performance—in the current study we followed Conway et al.'s (2010) usage and have referred to the learning measured by the task as ISL.

Viewing ISL as an outgrowth of memory systems also addresses the question of its domain generality. Rather than positing a unitary mechanism, recent accounts have viewed such learning as the outcome of basic memory processes, including activation, interference, integration, and chunking (Christiansen, 2019; Thiessen, 2017). Accordingly, statistical learning across modalities involves similar underlying mechanisms, but crucially these act on internal representations whose encoding is constrained by their modality among other features (Frost, Armstrong, Siegelman, & Christiansen, 2015). Results from studies implementing the memory game paradigm with stimuli from different input modalities have supported this view: Although comparable learning has been observed with elements of an artificial grammar mapped to visual, verbal, and multimodal stimuli (Conway et al., 2010; Karpicke & Pisoni, 2004), associations with language processing have been found only using stimuli that were presented auditorily (Conway et al., 2010) or stimuli that were visual but at the same time easy to encode verbally (Conway et al., 2007, 2010). In view of these findings, we have opted to use the auditory version of the task in which elements of the underlying artificial grammar are mapped to spoken color names.[3]

The memory game task indexed ISL as the improvement in recall for phonological sequences generated by an artificial grammar as well as baseline short-term memory for control (random) sequences. Our interest in PSTM arises due to its role in helping to establish the quality of the representations over which implicit learning mechanisms operate. Encoding the input is a prerequisite for both implicit and explicit learning; we do not claim that PSTM itself is implicit, but rather that it underpins the inductive processes at work in statistical learning. Although baseline PSTM has itself been associated with learning vocabulary and grammar in children and adults in both L1 and L2 (N. C. Ellis & Sinclair, 1996; French & O'Brien, 2008; Gathercole, Service, Hitch, Adams, & Martin, 1999; Service, 1992; Verhagen & Leseman, 2016; Verhagen, Leseman, & Messer, 2015), performance on memory based ISL tasks has now also been linked both to native language processing (Conway et al., 2010; Misyak & Christiansen, 2012; Misyak, Christiansen, & Tomblin, 2010) and language learning itself (Isbilen et al., 2018; Isbilen et al., 2017; Kidd, 2012).

The Present Study

The findings that we present here arose from a broad study designed to explore attainment in English by highly proficient, long-term daily users who had Polish as a L1, and whose exposure to English was in either an immersion learning context in the United Kingdom or a nonimmersion learning context in Poland. The study overall examined the influence of AoO and individual differences in PSTM and ISL ability. Foster, Bolibaugh, and Kotula (2013) previously reported on the results related to nativelike selection ability to identify which combinations of words are idiomatic in the L2 speech community in relation to AoO, exposure, motivation, and PSTM. The present study reports findings related to the grammatical attainment of the same long-term users of English. Using an aptitude treatment interaction design, we aimed to illuminate learning processes that are not directly observable but that are inferable from the interaction of cognitive abilities and AoO in different learning contexts. Specifically we ask two research questions:

- Research Question 1: To what extent does participants' performance on grammatical and ungrammatical items in a timed auditory GJT relate to their AoO and context of learning (immersed or nonimmersed)?
- Research Question 2: To what extent does participants' performance on grammatical and ungrammatical items in a timed auditory GJT relate to their PSTM and ISL ability, and how do these interact with AoO and context of learning (immersed and nonimmersed)?

Methods

Participants

The participants were 71[4] adult Polish-born speakers of English: 35 resided in West London in the United Kingdom, and 36 resided in Szczecin in Poland. Face-to-face interviews in both Polish and English, conducted by a trained International English Language Testing System instructor, determined that all were comfortably bilingual in both languages and were daily users of spoken and written English at the B2/C1 level in the Common European Framework of Reference for Languages.[5] All the participants were required to be current, daily users of English. For those living in Poland, this could be in any capacity, but, in most cases, it was for professional purposes. To avoid a criticism of previous studies (e.g., Johnson & Newport 1989) that involved users who had had only 5 years exposure to English and who had arguably not reached ultimate attainment, we included participants in our study who had had a minimum of 12 years' exposure to English, and, for many, the exposure had been for several decades. We could not be sure that 12 years is sufficient exposure, but DeKeyser (2012, p. 456) noted that there is empirical evidence for asymptote after 10 years of exposure for all dimensions of language "apart from vocabulary."

We measured AoO as the start of sustained and significant exposure and operationalized it as age of migration for the participants in the United Kingdom and as start of formal instruction for the participants in Poland. For the U.K. cohort, AoO ranged between one and 35 years ($M = 18$ years, $SD = 10$), whereas, for the Poland cohort, it was between 5 and 30 years ($M = 12$ years, $SD = 6$). The U.K. cohort was older at the time of test ($M = 56$ years, $SD = 17$) and had had longer exposure ($M = 38$ years, $SD = 19$, measured as their length of residence in the United Kingdom) relative to the Polish cohort (age at test: $M = 29$, $SD = 9$; exposure: $M = 16$ years, $SD = 5$). The participants in the U.K. cohort reported a mean of 2.65 years of English language study before arrival ($SD = 3.92$), although this learning varied in type and intensity, and 18 participants reported no formal instruction. We therefore characterized the learning of U.K. cohort as largely naturalistic, though these participants may have also had formal instruction in school or through other tuition. We also asked the participants to estimate their daily use of English on a 5-point scale ranging from 1 (*less than 25%*) to 5 (*more than 90%*). As one might have expected, the participants in the United Kingdom used more English than did those in Poland, estimating that nearly half of their daily language use was in English ($M = 2.85$, $SD = 1.18$), compared to an average of just over 25% for those living in Poland ($M = 1.47$, $SD = 0.77$). Given significant differences in age between the cohorts (and a negative association between age at test and PSTM in the U.K. cohort), we included age at test as a control predictor variable in all models.

In addition, we recruited 30 adult monolingual native speakers of English (i.e., people born in the United Kingdom) to give a baseline for the GJT. The native speakers were aged between late 20s and late 50s, all were resident in West London, and none had any experience of teaching or substantial formal study of grammar. Although a monolingual native speaker standard might not be considered appropriate in some areas of SLA research, it was important for our purposes to be sure that the baseline against which the participants' GJT performance was measured was not affected by the influence of any other language or indeed by any substantial study of the structure of English.

Research Instruments

We employed four instruments in the study that we administered in the following order: a combined measure of PSTM and ISL ability, a nativelike selection task (reported in Foster et al., 2013), a GJT, and a language use questionnaire (also reported in Foster et al., 2013). We gave the nonnative speaker participants all of the tests and only the GJT to the native speaker participants.

Phonological Short-Term Memory and Implicit Statistical Learning

We measured PSTM and ISL in a single serial recall task adapted from Karpicke and Pisoni (2004). We told the participants that they were going to undertake a memory test in which they would listen to and reproduce sequences. During the task (illustrated in Figure 1), the participants listened to a recorded voice in Polish—the female Polish voice Ewa from the Iwona™ voices of the TextAloud (2009) text-to-speech software—producing sequences made up of four color names: red, green, blue, and yellow. The sequences were four to eight items in length, with a 250-milliseconds interval between each item. The screen remained blank during the auditory presentation of the sequences. Following the final item in a sequence, a question mark appeared on the screen for 1,250 milliseconds as a cue that a response was required. The next screen displayed a four-colored grid. The assignment of color to grid position was randomly determined at the start of the experiment and then remained constant throughout the experiment. The participants repeated the sequences aloud while using a mouse to reproduce the sequence by clicking on the colored squares. Unbeknown to the participants, the sequences had been generated by an underlying finite-state grammar.[6] The participants were also unaware that, although the first 32 sequences were generated by the

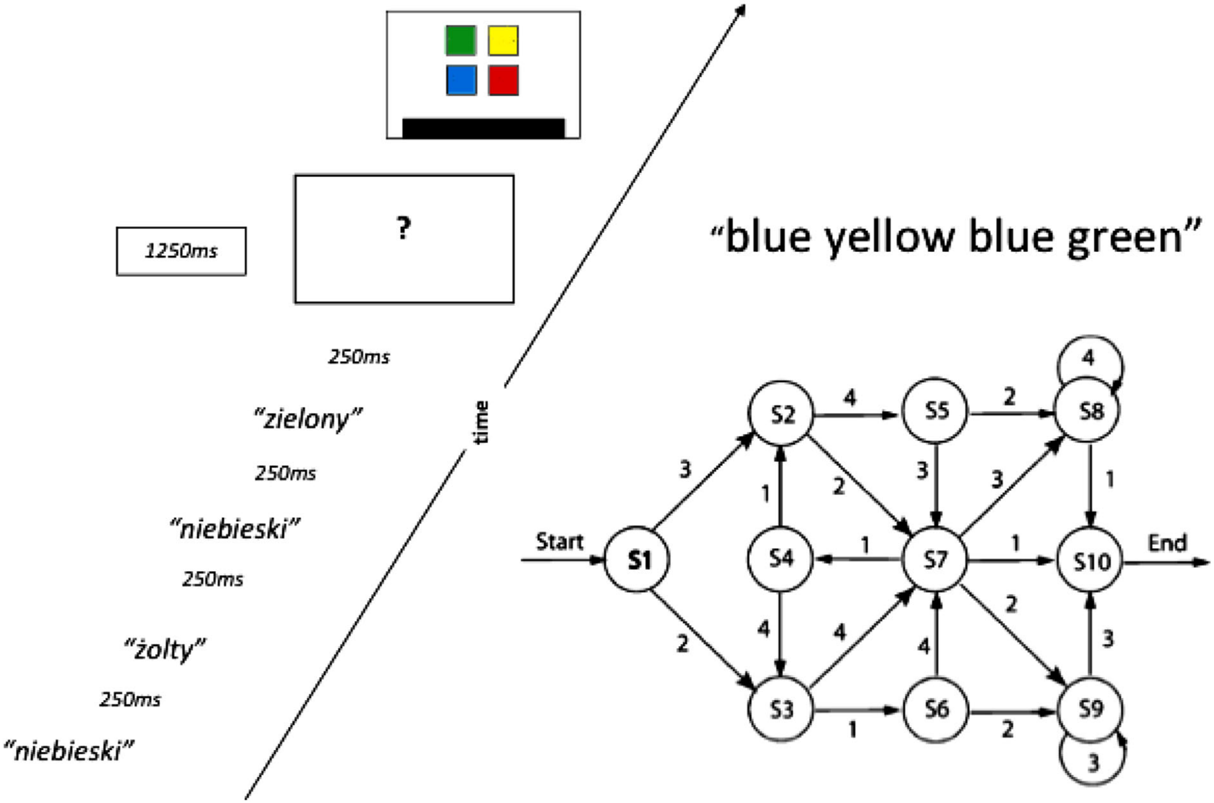

Figure 1 Illustration of phonological short-term memory and implicit statistical learning task. Training and test sequences were generated from Grammar A in Karpicke and Pisoni (2004), mapped to four color elements and auditorily presented in Polish. [Color figure can be viewed at wileyonlinelibrary.com]

grammar, the following 30 sequences—the test phase—were divided between novel grammatical and novel pseudorandom (ungrammatical) sequences. From the participants' point of view, there was no break or transition between the two phases, and the task (i.e., recalling sequences of varying length) remained the same throughout. Training and test sequences were generated from Grammar A used in Karpicke and Pisoni (2004) and in Conway et al. (2007), and the procedure followed the auditory condition of Karpicke and Pisoni (2004). We made two key modifications to the procedure: the presentation of the stimuli in Polish and the response mode (using a mouse instead of a button box). The testing and training sets generated by the grammar can be found in Conway et al. (2007).

We obtained two measures from the task: PSTM, which we measured as memory span for pseudorandom sequences during the test phase, and ISL, which we calculated as the improvement in memory span for the novel grammatical sequences compared to the pseudorandom (ungrammatical) sequences during the test phase. Following Conway et al.'s (2007) study, we used a weighted span scoring method in which we scored each correct sequence as the number of its constituent elements. We then summed these to arrive at a score for pseudorandom and grammatical test sequences for each participant.

Grammaticality Judgment Test

We developed our GJT measure from the 212-item GJT used by DeKeyser (2000). To reduce the possibility of loss of concentration in such a long test, we used 110 items.[7] To reduce the possibility of a ceiling effect, we based the items on 11 English structures most commonly mistaken by Polish L1 users of English that we garnered from informants who were experienced teachers of English in Poland. These structures were: *will* after adverbial, pronoun gender, present perfect, third person singular V + *-s*, V +*-ing* or infinitive, WH-question inversion, auxiliary verbs, definite and indefinite articles, plural on mass nouns, conditionals, and yes/no question inversion. Each structure was employed in 10 items, half grammatical and half ungrammatical, presented in a random order. All data elicitation materials from this study (Bolibaugh & Foster, 2021a) are available at both IRIS (https://iris-database.org) and OSF (https://doi.org/10.17605/OSF.IO/GCMXK). One male and one female speaker of English with London accents recorded the items to an audiofile. To help avoid loss of concentration by the participants, we alternated the speakers every 10 items. Each item was repeated after a one-second gap, and the next item followed after a further one-second gap. The whole test took about 40 minutes.

We instructed the participants to listen to the recording and respond to each item by marking it as grammatical or ungrammatical on an answer sheet. We told them that they could not stop the recording or ask for anything to be replayed or explained once the test began. The test proper began after five practice items.

Data Analyses

In order to simultaneously estimate the importance of predictors related to items and participants on GJT performance, we fit a series of mixed effects logistic regression models, using R (Version 3.5.1; R Core Team, 2018) and the R-packages lme4 (Version 1.1.21; Bates, Mächler, Bolker, & Walker, 2015), and lsmeans (Version 2.27.62; Lenth, 2016). The outcome variable for all models was GJT response accuracy, with each response coded as 1 if correct and 0 if incorrect. Predictors included sentence grammaticality (grammatical or ungrammatical), group (native speaker, nonnative speaker in the United Kingdom, or nonnative speaker in Poland), and four continuous variables: AoO, age at test, PSTM, and ISL. We fit models with crossed random effects for participants and items respectively, using a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). When models did not converge, we first simplified the random effects through removal of correlations between random slopes and intercepts and then by removal of slopes with the least variance. All models fit through this selection process can be found in the code supplement (Bolibaugh & Foster, 2021b) at both IRIS (https://iris-database.org) and OSF (https://doi.org/10.17605/OSF.IO/GCMXK).

We have reported the model results as odds and odd ratios, which are measures of effect size. An odds ratio of 1 indicates that the predictor was associated with no change in GJT response accuracy relative to baseline odds. Odds ratios of less than 1 indicate that the predictor was associated with a decrease in GJT accuracy, and odds ratios greater than 1 are an indication that the predictor was associated with an increase in GJT accuracy. We used an alpha level of .05 for all statistical tests.

We standardized all continuous predictors with mean of 0 and standard deviation of 1. Odds ratios for continuous predictors therefore indicate the change in GJT accuracy associated with a one standard deviation change in the predictor when all other continuous predictors were held at their mean. Because of the difficulty in interpreting the substantive size of an effect reported in odds ratios, we also used the inverse logit transformation to report probabilities as an illustration of model effects where appropriate. We treatment coded the categorical predictors with the reference category specified for all results.

We first report the descriptive statistics for all measures and then provide a series of analyses. Analysis 1 compared the effect of grammaticality in each of the three groups. In answer to Research Question 1, Analysis 2 explored how AoO interacted with grammaticality for the nonnative speakers only. In answer to Research Question 2, Analysis 3 explored separate models for the nonnative speakers in the United Kingdom and for the nonnative speakers in Poland, examining the effects of PSTM and ISL in relation to AoO and grammaticality.

Results

Descriptive Statistics

Tables 1 and 2 present descriptive statistics and bivariate correlations for all measured and observed variables. For native speakers, GJT accuracy ranged between 92% and 100%, with a mean of 96% ($SD = 2$). For nonnative speakers in Poland, accuracy ranged between 47% and 94%, with a mean of 74% ($SD = 14$) and Kuder-Richardson 20 reliability of .93. For the nonnative speakers in the United Kingdom, accuracy ranged between 53% and 94%, with a mean of 81% ($SD = 11$) and Kuder-Richardson 20 reliability of .91. On the PSTM measure, the mean recall of the nonnative speakers in Poland was 26.86 ($SD = 11.64$), whereas the mean recall score of the nonnative speakers in the United Kingdom was 18.38 ($SD = 10.75$). The mean ISL score for the Poland cohort was 7.17 ($SD = 10.8$), which was significantly different from zero ($p < .001$), indicating a group learning effect, but for the U.K. cohort it was 2.18 ($SD = 8.3$), which was not significantly different from zero ($p = .068$). In the U.K. cohort, there was no relationship between PSTM and ISL, $r = -.09$, $p > .05$, whereas ISL was unexpectedly negatively related to PSTM in the Poland cohort, $r = -.38$, $p < .05$.

These results may be the result of the different age profiles of the two nonnative speaker groups. As we noted earlier, the participants in the United Kingdom had a mean age at test of 56 years, but the participants in Poland were significantly younger, with a mean age at test of 29 years. We therefore investigated whether between group differences in the two cognitive predictors could be attributed to age. Differences between groups in ISL were significant, $b = 5.33$, 95% CI [1.06, 9.59], $t(76) = 2.48$, $p = .015$, but disappeared when we controlled for age at test, $b = 1.54$, 95% CI [−4.67, 7.74], $t(75) = 0.49$, $p = .623$. Similarly, differences between groups in PSTM were also significant, $b = 8.79$, 95% CI [3.84, 13.73], $t(76) = 3.54$, $p = .001$, but disappeared when we controlled for age at test, $b = -0.97$, 95% CI [−7.62, 5.68], $t(75) = -0.29$, $p = .772$. We therefore included age at test in all models with cognitive predictors.[8]

Table 1 Means, standard deviations, and Pearson correlations [with 95% confidence intervals] for all outcome and predictor variables for nonnative speakers in the United Kingdom ($n = 35$)

Variable	*M*	*SD*	1	2	3	4	5
1. GJT	0.81 [0.77, 0.85]	0.11	—				
2. Age	55.91 [49.98, 61.84]	17.26	−.23 [−.52, .11]	—			
3. AoO	18.17 [14.78, 21.56]	9.86	**−.71** [−.84, −.49]	.13 [−.21, .44]	—		
4. LoE	37.74 [31.31, 44.18]	18.74	.16 [−.18, .47]	**.85** [.73, .92]	**−.41** [−.65, −.09]	—	
5. PSTM	18.38 [14.69, 22.07]	10.75	**.36** [.03, .63]	**−.69** [−.83, −.46]	−.15 [−.47, .20]	**−.55** [−.75, −.26]	—
6. ISL	2.18 [−0.67, 5.02]	8.30	.18 [−.17, .48]	−.27 [−.55, .08]	−.26 [−.55, .09]	−.10 [−.43, .24]	−.09 [−.42, .25]

Note. Values of r in boldface are statistically significant at alpha = .05; exact p values are reported in Table S1.1 in Appendix S1 in the online Supporting Information. GJT = grammaticality judgment test; AoO = age of onset; LoE = length of exposure; PSTM = phonological short-term memory; ISL = implicit statistical learning.

Table 2 Means, standard deviations, and Pearson correlations [with 95% confidence intervals] for all outcome and predictor variables for nonnative speakers in Poland (n = 36)

Variable	*M*	*SD*	1	2	3	4	5
1. GJT	0.74	0.14	—				
	[0.70, 0.79]						
2. Age	28.89	9.29	−.29	—			
	[25.75, 32.03]		[−.57, .04]				
3. AoO	12.17	5.87	**−.53**	**.82**	—		
	[10.18, 14.15]		[−.73, −.24]	[.67, .90]			
4. LoE	16.31	4.97	.10	**.73**	.25	—	
	[14.62, 17.98]		[−.24, .41]	[.53, .86]	[−.08, .54]		
5. PSTM	26.86	11.64	−.02	−.05	−.06	.02	—
	[22.92, 30.80]		[−.34, .31]	[−.37, .29]	[−.38, .27]	[−.31, .34]	
6. ISL	7.17	10.85	.09	−.16	−.19	−.07	**−.38**
	[3.50, 10.84]		[−.25, .41]	[−.46, .18]	[−.48, .15]	[−.39, .27]	[−.63, −.06]

Note. Values of r in boldface are statistically significant at alpha = .05; exact p values are reported in Table S1.2 in Appendix S1 in the online Supporting Information. GJT = grammaticality judgment test; AoO = age of onset; LoE = length of exposure; PSTM = phonological short-term memory; ISL = implicit statistical learning.

Analysis 1: Grammaticality by Group

We first examined GJT performance in relation to grammaticality across groups (see Figure 2; R model statement for Analysis 1: Response_Accuracy $\sim$ Group*Grammaticality + (1|Subject) + (1|Sentence), family = 'binomial'). There was a main effect of group, $\chi^2(2) = 84.56$, $p < .001$, because both Polish L1 groups were less accurate than native speakers. There was also a main effect of grammaticality, $\chi^2(1) = 12.3$, $p < .001$, because all groups were less accurate when judging ungrammatical sentences. These effects were qualified by an interaction of group and grammaticality, $\chi^2(2) = 6.29$, $p < .044$. Whereas the decrease in performance on ungrammatical sentences was of a similar magnitude for the native speakers and the nonnative speaker participants in the United Kingdom, $b = 0.33$, $SE = 0.23$, 95% CI [−0.11, 0.77], $z = 1.46$, $p = .144$, the magnitude of the decrease in performance on ungrammatical sentences was significantly larger for the nonnative speaker participants in Poland than for native speakers, $b = 0.52$, $SE = 0.22$, 95% CI [0.08, 0.96], $z = 2.33$, $p = .019$. Post hoc comparisons using the Tukey HSD test revealed that the U.K. cohort was significantly better at recognizing grammatical sentences ($M = 0.86$, $SD = 0.35$) than was the Poland cohort ($M = 0.79$, $SD = 0.41$; $p = .019$). In contrast, correct rejections of ungrammatical sentences did not differ significantly between groups (U.K. cohort: $M = 0.76$, $SD = 0.43$; Poland cohort: $M = 0.70$, $SD = 0.46$; $p = .156$). Taken together these results highlighted a smaller grammaticality effect (better performance on grammatical than ungrammatical sentences) in the Poland cohort relative to the U.K cohort due to the Poland participants' greater likelihood of mistakenly rejecting grammatical sentences.

Analysis 2: Age of Onset and Grammaticality in Nonnative Speakers

We next explored the relationship between AoO and GJT performance in the L2 groups. GJT performance was strongly negatively associated with AoO for the nonnative speakers in the United Kingdom, $r(33) = -.71$, $p < .001$, and moderately negatively associated with AoO for the nonnative speakers in Poland, $r(34) = -.53$, $p < .001$. As the scatterplots in Figure 3 show, the advantage of early onset for the U.K. cohort was categorical because no participants with an AoO younger than 9 years scored lower than the lowest native speaker score, but this was not the case for the Poland cohort, for whom an early start was associated with a smaller relative advantage.

Table 3 reports a second mixed effects logistic regression model with group (nonnative speakers in the United Kingdom, nonnative speakers in Poland), grammaticality, and AoO (scaled) as predictors (R model statement for

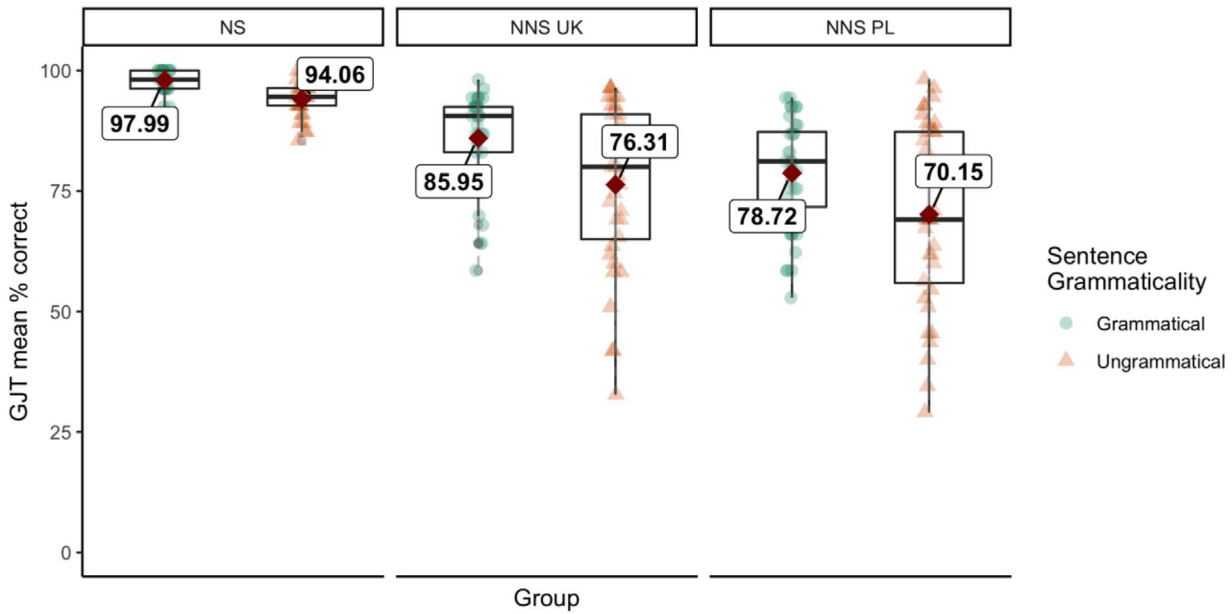

Figure 2 Grammaticality judgment test (GJT) mean percent correct by group for grammatical and ungrammatical sentences. NS = native speakers; NNS UK = nonnative speakers in the United Kingdom; NNS PL = nonnative speakers in Poland. [Color figure can be viewed at wileyonlinelibrary.com]

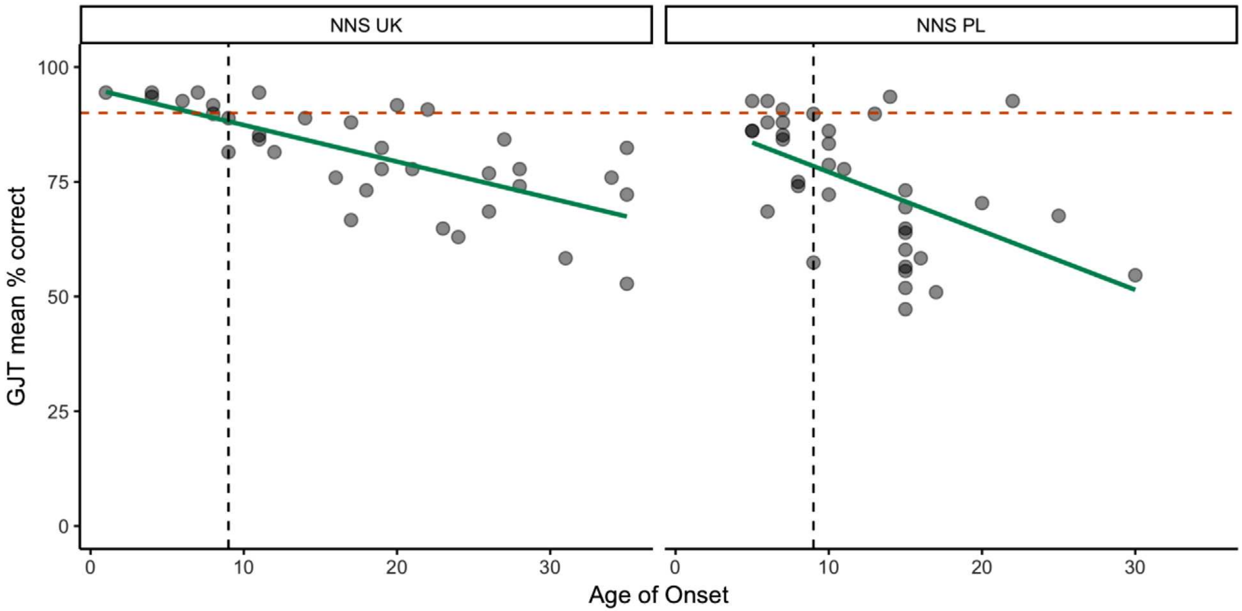

Figure 3 Scatterplots between grammaticality judgment test (GJT) mean percent correct and age of onset (in years) for nonnative speakers. The dashed horizontal line marks the lowest native speaker score. The dashed vertical line delimits the age of onset before which all U.K. nonnative speaker participants performed in the native speaker range. NNS UK = nonnative speakers in the United Kingdom; NNS PL = nonnative speakers in Poland. [Color figure can be viewed at wileyonlinelibrary.com]

Table 3 Odds of correct grammaticality judgment test responses by the predictors group, sentence grammaticality, age of onset, and their interactions

Predictors	*OR*	95% CI	*p*
Intercept	9.50	[6.58, 13.71]	<.001
NNS PL	0.44	[0.30, 0.64]	<.001
Ungrammatical	0.65	[0.44, 0.97]	.035
AoO	0.74	[0.59, 0.93]	.011
NNS PL × Ungrammatical	0.88	[0.67, 1.16]	.373
NNS PL × AoO	0.67	[0.44, 1.04]	.073
Ungrammatical × AoO	0.64	[0.54, 0.75]	<.001
NNS PL × Ungrammatical × AoO	1.58	[1.18, 2.11]	.002

Note. All continuous predictors were centered and scaled. The intercept represents the odds of a correct response when reading a grammatical sentence for a nonnative speaker participant from the United Kingdom with the mean age of onset. NNS PL = nonnative speakers in Poland; AoO = age of onset.

Analysis 2: Response_Accuracy ~ Group * Grammaticality * AoO_cs + (1 | Subject) + (1 |Sentence), glmerControl(optimizer = "bobyqa"), family = 'binomial'). In addition to the previously reported effects of group and grammaticality, we also observed a three-way interaction with AoO, $\chi^2(3) = 26.88$, $p < .001$, that can be seen in the plot of fitted values in Figure 4.

Immersed learners' judgment of *grammatical* sentences was relatively unaffected by starting age, with a small reduction in accuracy associated with each one standard deviation increase in AoO. With an AoO of 7 years (−1 *SD*), the nonnative speaker participants in the United Kingdom had a mean recognition accuracy of 93% for grammatical sentences, which decreased to a mean of 90% with an AoO of 15 years and to a mean of 88% with an AoO of 24 years (+1 *SD*). In contrast, the U.K. cohort's overall poorer performance on *ungrammatical* sentences was largely explained by the much greater reduction in accuracy associated with a later starting age (from a mean of 93% correct rejection with an AoO of 7 years to a mean of 86% with an AoO of 15 years and down to a mean of 75% with an AoO of 24 years). Thus, for those participants who had acquired their L2 as immersed immigrants, the grammaticality effect was larger for those with later onset, and overall age effects were largely driven by the lower performance on ungrammatical sentences.

In contrast to immersed learners, the nonnative speaker participants in Poland demonstrated additive effects of grammaticality and AoO: They became progressively more likely to mistakenly reject *grammatical* sentences as

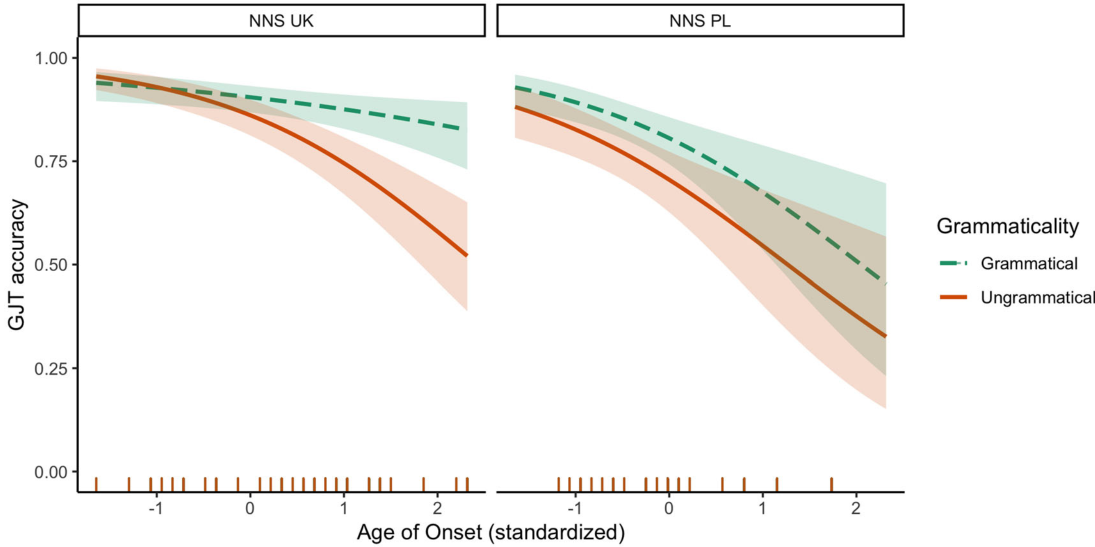

Figure 4 Predicted grammatical judgment test (GJT) accuracy for grammatical and ungrammatical sentences as a function of age of onset (standardized) and group. Age of onset is presented in standard deviation units, (−1 *SD* = 7 years age of onset, 0 *SD* = 15 years age of onset, and 1 *SD* = 24 years age of onset). Ribbons represent 95% confidence intervals. NNS UK = nonnative speakers in the United Kingdom; NNS PL = nonnative speakers in Poland. [Color figure can be viewed at wileyonlinelibrary.com]

starting age increased, *and* this decrease in accuracy was mirrored for ungrammatical sentences. Thus, with an AoO of 7 years (-1 *SD*), the Poland cohort accurately endorsed a mean of 89% of grammatical sentences and correctly rejected a mean of 83% of ungrammatical sentences. For a learner who started learning at 24 years ($+1$ *SD*), accuracy decreased to 67% for grammatical and to 55% for ungrammatical sentences.

Analysis 3: Phonological Short-Term Memory and Implicit Statistical Learning in Relation to Age of Onset and to Grammaticality

We next examined the role of our two cognitive predictors in each nonnative speaker group separately. We rescaled the continuous predictors (age at test, AoO, PSTM, and ISL) with reference to each nonnative speaker group; model coefficients therefore represent the GJT change associated with one standard deviation change in the predictor for that group. For each group, we started by fitting a model that included both the interaction of PSTM with AoO and grammaticality and of ISL with AoO and grammaticality. We evaluated whether interactions were justified by using likelihood ratio tests and the Akaike information criterion to compare nested models and progressively reduced each interaction term until a comparison with a reduced model signaled a worse fit. We retained all predictors of interest regardless of their significance, and the full suite of models and model comparisons (Bolibaugh & Foster, 2021c) can be seen online at both IRIS (https://iris-database.org) and OSF (https://doi.org/10.17605/OSF.IO/GCMXK). The best fitting model for each group is presented in Table 4 (R model statements for Analyses 3: Nonnative speakers in the United Kingdom: Response_Accuracy ~ Age_cs + Grammaticality*AoO_cs*pSTM_cs + Grammaticality*ISL_cs + (1|Subject) + (1|Sentence), control = glmerControl(optimizer = “bobyqa”), family = “binomial”. Nonnative speakers in Poland: Response_Accuracy ~ Age_cs + Grammaticality + AoO_cs + pSTM_cs + ISL_cs + (1|Subject) + (1|Sentence), control = glmerControl(optimizer = “bobyqa”), family = “binomial”).

The best fitting model for the U.K. cohort included a three-way interaction of grammaticality, AoO, and PSTM, $\chi^2(1) = 6.88$, $p < .01$, and a two-way interaction of grammaticality and ISL, $\chi^2(1) = 5.46$, $p < .02$. Higher scores for both PSTM and ISL were associated with higher accuracy on ungrammatical, but not grammatical, sentences. For PSTM, this relationship was further mediated by AoO; higher levels of PSTM grew more important in attenuating the drop in accuracy associated with ungrammatical sentences as AoO increased.

In the best fitting model for the Poland cohort, only the additive effects of AoO and grammaticality were significant (age at test, PSTM, and ISL were

Table 4 Final models by group for odds of correct grammaticality judgment test responses as a function of the predictors grammaticality, age at test, age of onset (AoO), phonological short-term memory (PSTM), implicit statistical learning (ISL), and their interactions

	NNS UK			NNS PL		
Predictors	*OR*	95% CI	*p*	*OR*	95% CI	*p*
Intercept	9.87	[6.59, 14.78]	**<.001**	5.23	[3.73, 7.34]	**<.001**
Age at test	0.74	[0.53, 1.04]	.084	1.46	[0.95, 2.24]	.088
Ungrammatical	0.67	[0.40, 1.12]	.131	0.56	[0.40, 0.78]	**.001**
AoO	0.68	[0.54, 0.86]	**.001**	0.46	[0.30, 0.71]	**.001**
PSTM	1.02	[0.74, 1.39]	.914	0.94	[0.72, 1.24]	.679
ISL	0.80	[0.62, 1.05]	.106	1.01	[0.77, 1.34]	.929
Ungrammatical × AoO	0.60	[0.48, 0.75]	**<.001**			
Ungrammatical × PSTM	1.07	[0.87, 1.32]	.515			
AoO × PSTM	1.01	[0.77, 1.33]	.921			
Ungrammatical × ISL	1.31	[1.05, 1.64]	**.017**			
Ungrammatical × AoO × PSTM	1.41	[1.10, 1.81]	**.008**			

Note. Values of p in boldface are statistically significant at alpha = .05. All continuous predictors were centered and scaled with reference to either nonnative speakers in the United Kingdom or in Poland. The intercepts represent the odds of correct response when reading a grammatical sentence, for a participant with mean age at test, mean AoO, mean PSTM, and mean ISL. NNS UK = nonnative speakers in the United Kingdom; NNS PL = nonnative speakers in Poland.

retained for comparison to the U.K. cohort). Neither cognitive predictor was associated with GJT accuracy, with both odds ratios very close to 1.

Summary of Findings

In summary, we found that all the participants, monolingual and bilinguals, were less accurate on ungrammatical sentences relative to grammatical sentences. For immersed learners (the U.K. cohort), the magnitude of this difference was not distinguishable from the native speakers, even though they were less accurate overall. The nonnative speaker participants in the United Kingdom were relatively accurate when judging grammatical sentences, regardless of their starting age, whereas age effects were most pronounced for correct rejection of ungrammatical items (i.e., the lower the AoO, the more likely they were to correctly reject ungrammaticalities). Better ISL ability mitigated poorer performance on ungrammatical sentences for all the participants in the U.K. cohort, and PSTM similarly facilitated correct rejection of ungrammatical sentences but interacted with AoO such that it was most important for those with a later start.

The participants in Poland showed a different pattern of performance. They were less accurate when judging grammatical sentences than the U.K. cohort, and demonstrated marginally stronger AoO effects than the U.K. cohort for grammatical sentences as well. The fact that later starters in the Poland cohort were more likely to mistakenly mark grammatical sentences as ungrammatical thus appears responsible for the U.K.–Poland cohort difference in accuracy on grammatical sentences. The negative association between AoO and GJT accuracy was roughly the same for both grammatical and ungrammatical sentences. In further contrast to the immersed learners, neither PSTM nor ISL was associated with GJT performance in the Poland cohort.

Discussion

The aim of the present study was to assess how far, and in what conditions, ISL mechanisms influence morphosyntactic attainment in immersed and instructed learners. Our first research question asked about the extent to which participants' performance on grammatical and ungrammatical items in a timed auditory GJT related to their AoO and context of learning (age by context). The second research question asked about the extent to which performance was related to measures of PSTM and ISL ability relative to AoO (age by aptitude). We contextualize our findings for each interaction before offering a possible account for the complex pattern of findings.

Age by Context Interactions

In line with results from nearly all previous studies of age effects in immersed learners, we found a negative relationship between starting age and ultimate attainment. The magnitude of the relationship was also consistent with previous studies, highlighting the robustness of the effect. Slightly more surprisingly, we also found a negative, statistical association between AoO and GJT accuracy in our cohort living in Poland, who had started as instructed learners and eventually become daily users of their L2. Our findings of a medium-sized effect are similar to the findings reported by Larson-Hall (2008), who also used an auditory GJT, but with long-term Japanese learners of English in Japan.

We do not interpret these findings to mean that AoO played the same role in both learning contexts. Age effects in immersed learners were categorical, in that an early start (AoO $<$ 9 years) guaranteed performance indistinguishable from monolingual English speakers. This contrasted with the relative advantage conferred by an early start in an instructed context where many early-onset learners performed outside the monolingual range. Thus, early onset benefitted performance in both contexts, but only in immersed learners was it sufficient to guarantee performance in the monolingual range.

Considering the role of grammaticality gives further insight into group differences. Only in the immersed group did AoO influence the magnitude of the grammaticality effect. We find it interesting that the grammaticality effect in adult-onset immersed learners' performance thus resembled that of adult participants in laboratory studies of incidental or implicit learning of morphosyntax, who also failed to learn to reject ungrammatical strings (Andringa & Curcic, 2015; Grey, Williams, & Rebuschat, 2015; Rebuschat & Williams, 2012). This commonality arose despite large differences in the amount of input received in these two types of studies (i.e., most laboratory studies offer fewer than 10 hours of exposure, whereas our participants had an average length of exposure of 38 years). In contrast, the grammaticality effect in instructed learners was independent of AoO effects. Thus, early-onset learners in Poland not only did better relative to later starters, but this advantage combined with the grammaticality effect meant that early-onset instructed learners performed near ceiling on grammatical sentences but were less consistent as a group on ungrammatical sentences.

Age by Aptitude Interactions

Our cognitive predictors played no role in the GJT performance of instructed learners. These findings are similar to those reported in Suzuki and DeKeyser's (2017) study of advanced Chinese L1 learners of L2 Japanese: Neither a letter span task nor a serial reaction task correlated with scores on timed GJTs (or with other online measures). Our results thus extend Suzuki and DeKeyser's null effects for short-term memory and implicit learning in adult-onset learners to early-onset instructed learners as well.

In immersed learners, our cognitive predictors were both associated with GJT performance. ISL predicted less poor performance on ungrammatical sentences regardless of AoO, and PSTM also predicted improved performance on ungrammatical sentences but grew in importance as AoO increased. Two conclusions can be drawn. The first confirms that statistical learning ability is associated with learning from naturalistic input, thus pointing to data driven, or inductive, learning processes. Our results are consistent with Granena's (2013) findings in immersed Chinese L1 learners of L2 Spanish: Performance on two sequence learning measures, LLAMA D and serial reaction time, correlated with agreement error knowledge in early and late learners, respectively. The second implication is that, counter to the fundamental difference hypothesis, our findings suggest continuity in learning processes between child and adult learners. These results parallel those of Abrahamsson and Hyltenstam (2008), who found that performance on the LLAMA aptitude battery predicted linguistic attainment in both early- and late-onset learners. We elaborate on these conclusions below.

How Far and in What Conditions Is Implicit Learning Preserved in Second Language Acquisition?

Following the logic of DeKeyser (2012, among others), we had suggested that associations between cognitive variables and linguistic knowledge can be used to make inferences about the learning processes underlying the acquisition of that knowledge. Thus, using PSTM and ISL as diagnostic criteria for data-driven, inductive, and possibly implicit, learning, would lead us to conclude that our participants in instructed and immersed contexts engaged different learning mechanisms. This conclusion is in line with the view that initial provision of explicit and metalinguistic input likely reduces reliance on chunking as a learning mechanism in foreign learning contexts, regardless of the length of subsequent exposure and use. Our early-onset nonnative-speaker participants in Poland were less likely than later-onset learners in Poland to misidentify grammatical sentences as ungrammatical; this may therefore indicate automatized explicit knowledge. A greater reliance on explicit knowledge, which is fragmentary and anomalous (R. Ellis, 2004), is consistent with the performance of our later-onset instructed learners who were more likely to misidentify grammatical sentences.

What about our immersed learners? Rather than indicate either a qualitative shift in processing or a quantitative decline in general implicit learning ability (Granena, 2013; Granena & Long, 2013), we suggest that the AoO function for GJT performance in immersed learners reflects a decline in the ability to learn from implicit negative evidence, that is, to infer which patterns are *not* grammatical as a function of exposure to positive evidence only, and that this decline can be explained by an account that posits continuity in learning processes between children and adults.

The question of how children come to learn what is not part of their language merely from exposure to what is part of it has long been a central point of contention in language learning theory (Clark & Lappin, 2010; Goldberg, 2003; Gold, 1967; Miller & Chomsky, 1963). How can an individual know that a certain pattern is not licensed rather than simply not yet encountered? Empiricist theories overcome this logical problem of language acquisition by representing linguistic knowledge as the outcome of a stochastic process, where the statistical structure of an individual's experience enables probabilistic inference: "[e]ssentially, if a particular grammatical construction is not observed during some extended but finite exposure, one can safely assume that it is not part of the language" (Rohde & Plaut, 1999, p. 3). In the absence of explicit negative feedback, knowing that a given utterance is ungrammatical thus requires learners, first, to realize that they have rarely (if ever) encountered it before and, second, to realize that, if it were licensed, they likely would have encountered it (Clark & Lappin 2010). In this account, statistical underrepresentation, and statistical preemption, are what gives rise to native speaker intuitions regarding violations of grammaticality. For learners to derive intuitions about what is not part of the language, they require probabilistic evidence about what is in the language from a high fidelity, representative sample.

Memory-based perspectives on ISL, which highlight the role of encoding and integration, and draw on research concerning prototype formation in exemplar models (Christiansen, 2019; Thiessen, 2017) can help to explain why, in our data, grammaticality interacted with AoO and with our cognitive predictors. When speakers encounter a grammatical sentence in a judgment task, they need only a fuzzy prototype of that construction to serve as a match in order for them to give a correct endorsement. When speakers encounter an ungrammatical sentence, however, this same prototype may also be a match if it is not sufficiently detailed to enable them to discriminate it from a grammatical sentence. Thus, if "statistical underrepresentation must do the job of negative evidence" (Pierrehumbert, 2003, p. 196) in data-driven learning, it is not surprising that the participants who used statistical evidence more efficiently in our ISL task were better able to reject ungrammatical sentences in the GJT.

The three-way interaction of grammaticality, AoO, and PSTM further supports this explanation. Developmental changes in encoding (whereby children encode seemingly irrelevant details of the input including speaker specific information; Houston & Jusczyk, 2000) and in cognitive control (Thompson-Schill, Ramscar, & Chrysikou, 2009), ally to entrench the perceptual salience of L1 regularities as AoO increases. Because the ability to encode L2 input in a manner that precisely represents the underlying linguistic regularities diminishes with age, so too then does the ability to detect when ungrammatical sentences are statistically underrepresented. Because PSTM is responsible for the quality of the representations from which statistics are abstracted, it is again unsurprising that those participants with better PSTM, presumably equipped with a higher fidelity language sample, were better able to detect ungrammatical sentences. Other reasons for a grammaticality effect in immersed learners may be put forward—learners' tendency to accept ungrammatical sentences may simply result from a response bias toward answering yes. Smaller associations between cognitive predictors and accuracy on grammatical sentences might then emerge from a truncated range of scores for grammatical sentences in older-onset learners. However, this account does not explain why such a response bias would be larger in those with later onset.

Finally, it is important to note that our account does not question the importance of high verbal analytic ability in later-onset learners; rather we suggest that the effects of explicit aptitude reported in previous studies should not be interpreted to entail a general loss of implicit learning ability.

Grammaticality Judgment Tests as Measures of Implicit and Explicit Knowledge

In the account detailed above, we proposed that, for our immersed learners, AoO-dependent increases in the magnitude of the grammaticality effect were consistent with probabilistic judgments that were informed by an implicitly acquired knowledge base. PSTM and ISL were associated with performance on ungrammatical sentences because task demands required more precise representations. In contrast, our instructed cohort was more variable when judging grammatical sentences and became progressively more likely to mistakenly reject grammatical sentences at later starting ages. This is consistent with the application of explicit, rule-based knowledge, a conclusion supported by the lack of association between GJT performance and our cognitive predictors.

Taken together, these conclusions highlight the importance of considering how the history of learners shapes the representations that they bring to the task.

Measuring Implicit Statistical Learning

We used a serial recall-based memory task as a measure of implicit learning. Our task worked well enough, but certain limitations were apparent. Group differences in both PSTM and ISL measures were related to age at test. Relationships with age were more apparent in our U.K. cohort due to the fact that these participants were significantly older than the Poland cohort. We also found a negative relationship between baseline PSTM and ISL in our Poland cohort. In other words, those participants with the highest baseline PSTM improved the least on the learning task. We assume that this is an artefact of the task that had an upper length of eight elements for each sequence. Although the task had been normed on undergraduates in the United States by Karpicke and Pisoni (2004) as well as by Conway et al. (2007), it is possible that our relatively highly educated participants in Poland (most with postgraduate degrees) were performing near ceiling. This attenuation of the upper range in the nonnative speaker group in Poland means the lack of correlation between ISL and GJT performance in this Poland cohort will need to be verified. Further limitations include a lack of reliability measures for the task. Although Isbilen et al. (2017) found that statistically induced chunking recall tasks fare much better in test-retest reliability than traditional judgment-based statistical learning tasks, difference scores (i.e., ISL is calculated as the difference between performance on statistically regular and statistically irregular patterns) are known to be less reliable. In the present study, the dependence between learning trials means there was no straightforward way of calculating internal consistency or split-half reliability in a single test administration, and test-retest reliability will need to be established. Finally, although Karpicke and Pisoni (2004) found no evidence of the development of verbalizable knowledge and although Christiansen and colleagues have argued that the particular features of serial recall-based implicit learning tasks preclude the use of strategic or explicit processing (Christiansen, 2019; Isbilen et al., 2017), we did not administer measures of awareness, and the participants' use of explicit knowledge in the task cannot be ruled out in the present study.

Limitations and Directions for Future Research

These findings, although suggestive, result from an exploratory study. The complex interactions will need to be replicated in larger samples to verify the roles of our cognitive predictors and their interactions with age and context.

Any confirmatory models will need to be sufficiently powered to support a full random effects structure. Our data and code are available for future syntheses using participant level data.

An important question will be to determine the extent to which the findings can be generalized to other samples. The demographic profile of our sample was strongly influenced by sociopolitical factors in both the United Kingdom (where immigration from Poland peaked during and immediately after the Second World War and following European accession in 2004), and in Poland (where very few people studied English prior to the fall of the Communist government in 1989). Those of our participants who did so were therefore unique in different ways. In the United Kingdom, the participants who started earlier tended to be older and have had longer exposure. In Poland, most participants had started learning English after 1989, and thus were younger.[9] Although our participants were screened to ensure that they were fluent Polish users (all participants reported a minimum of 25% daily use of Polish) as well as having comparable oral proficiency in English within and across groups, it is clear that language dominance varied both between participants and also within participants across time. This may have had implications for the automaticity of lexical retrieval in the memory task.

The question of whether individual differences in a cognitive ability or an aptitude can be measured years after learning has occurred is also an important one; susceptibility to training, experience, or age would place limitations on the assumptions that can be made regarding an aptitude's involvement in learning that has already happened. Our two cohorts differed significantly in mean age at test, as well as in their performance on the cognitive measures. We found that between-group differences in PSTM and ISL disappeared when age at test was considered, and therefore we included age at test as a control variable in all subsequent models, but to verify our conclusions, the study will need to be replicated in a sample in which multicollinearity is physically controlled through sampling. Finally, although the measurement of statistical learning as improvement in memory span is promising, a valuable direction for future study will be to determine the reliability of the measures as well as the nature of any relationship between baseline PSTM and statistical learning.

Conclusion

This study explored the interactions of AoO and cognitive abilities hypothesized to underpin implicit learning in order to infer the extent of the influence of implicit learning mechanisms on the ultimate morphosyntactic knowledge of long-term second language speakers. We examined these interactions in speakers who had acquired their L2 either in a foreign language context or as immigrants in an immersed language context. Our findings confirm that, even though all our Polish L1 participants were long-term, functional bilinguals who used their L2 on a daily basis, the context in which they had learned their L2 shaped their morphosyntactic attainment: Group differences in the interaction of AoO with grammaticality (i.e., grammatical vs. ungrammatical sentences) and in the role played by PSTM and ISL suggest that qualitatively different types of knowledge can underlie superficially similar task performance.

We observed a growing dissociation between grammatical and ungrammatical performance (a grammaticality effect) when onset of learning was delayed in immersed learners; this effect was attenuated in those with better PSTM and statistical learning abilities. We propose that these results reflect a decline in the ability to learn from implicit negative evidence that emerges naturally from a continuous, statistically based learning mechanism operating over a progressively lower fidelity language sample.

In a foreign language context, GJT performance was subject to additive effects of AoO and grammaticality and was not associated with either cognitive predictor, suggesting that implicit learning mechanisms were not involved in the development of the knowledge assessed by the GJT.

Final revised version accepted 20 December 2020

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://doi.org/10.17605/OSF.IO/Z6QUW; https://doi.org/10.17605/OSF.IO/J3DK6 and http://www.iris-database.org. All proprietary materials have been precisely identified in the manuscript.

Notes

1 A reviewer noted that competence in a Chomskyan sense is implicit but declarative.

2 It is generally agreed that implicit knowledge is an outcome of implicit learning, though not the only possible outcome—learners may spontaneously develop (explicit) insight into their own facilitated processing routines, a process that Karmiloff-Smith described as representational redescription (Clark & Karmiloff-Smith, 1993; Karmiloff-Smith, 1992).

3 A reviewer noted that using a verbal measure of ISL raises the concern that whenever correlations arise between it and a language measure, it is simply due to measuring the same construct (e.g., verbal ability) twice. In the present study, we adopted the view that ISL is a domain general mechanism constrained by the representational code over which it operates.

4 Foster, Bolibaugh, and Kotula (2013) reported on 79 participants. For the particular analyses conducted here on the dataset, eight participants had some missing data and were therefore excluded.

5 The Common European Framework of Reference for Languages was established by the Council of Europe to standardize the levels of language exams. Level C1 describes effective operational proficiency in terms of appropriateness, sensitivity, and capacity to deal with unfamiliar topics and may be equated to an advanced level user. Level B2 is just below Level C1.

6 The arrows in the grammar define the permissible transitions between elements to generate sequences of varying lengths. By beginning at the Start node and following the arrows through the different nodes, the grammar generates sequences of varying lengths, with elements ordered according to the same underlying rules. The four constituent elements (numbers 1–4) were randomly assigned to four color names: 1 (yellow), 2 (blue), 3 (green), and 4 (red). Thus, the example sequence 2-1-2-3 would become blue-yellow-blue-green. Unlike in Conway et al.'s (2007) study, the sequences were auditorily presented in Polish (the L1 of all bilingual participants).

7 We found that one of the sentences had been duplicated. We therefore ran all analyses on 108 items.

8 In contrast to AoO, length of exposure was not significantly associated with GJT performance in either group. Given the linear dependencies between age at test, AoO, and length of exposure in the U.K. cohort, we did not include length of exposure in further analyses, but we included age at test in all analyses that included a cognitive predictor.

9 Many of these younger participants in Poland also started learning English at the same age, 13 years, due to curricular reforms. A reviewer noted that the wide range of performance associated with this AoO group may have led to model overfitting.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481–509. https://doi.org/10.1017/S027226310808073X

Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing: Evidence from differential object marking acquisition. *Studies in Second Language Acquisition*, *37*, 237–268. https://doi.org/10.1017/S0272263115000017

Baumert, J., Fleckenstein, J., Leucht, M., Köller, O., & Möller, J. (2020). The long-term proficiency of early, middle, and late starters learning English as a foreign language at school: A narrative review and empirical study. *Language Learning*, *70*, 1091–1135. https://doi.org/10.1111/lang.12414

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning*, *29*, 81–103. https://doi.org/10.1111/j.1467-1770.1979.tb01053.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*, 3–49.

Bolibaugh, C., & Foster, P. (2021a). *Grammatical Judgement Task. Materials from "Implicit statistical learning in naturalistic and instructed morphosyntactic attainment: An aptitude-treatment interaction design"* [Language test]. IRIS Database, University of York, UK. https://doi.org/10.48316/vyme-a579

Bolibaugh, C., & Foster, P. (2021b). *R code. Materials from "Implicit statistical learning in naturalistic and instructed morphosyntactic attainment: An aptitude-treatment interaction design"* [Software/Analysis code]. IRIS Database, University of York, UK. https://doi.org/10.48316/dsc8-ys03

Bolibaugh, C., & Foster, P. (2021c). *Datasets from "Implicit statistical learning in naturalistic and instructed morphosyntactic attainment: An aptitude-treatment interaction design"* IRIS Database, University of York, UK. https://doi.org/10.48316/91ds-4g40

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*, 247–271. https://doi.org/10.1017/S0272263110000756

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*, 455–586. https://doi.org/10.1111/tops.12332

Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind & Language*, *8*, 487–519. https://doi.org/10.1111/j.1468-0017.1993.tb00299.x

Clark, A., & Lappin, S. (2010). *Linguistic nativism and the poverty of the stimulus*. Hoboken, NJ: Wiley. https://doi.org/10.1002/9781444390568

Cleary, M., Pisoni, D. B., & Geers, A. E. (2001). Some measures of verbal and spatial working memory in eight- and nine-year-old hearing-impaired children with cochlear implants. *Ear and Hearing*, *22*, 395–411. https://doi.org/10.1097/00003446-200110000-00004

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*, 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*, 356–371. https://doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*, 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Conway, C. M., Karpicke, J., & Pisoni, D. B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of Deaf Studies and Deaf Education*, *12*, 317–334. https://doi.org/10.1093/deafed/enm019

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533. https://doi.org/10.1017/s0272263100004022

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, *63*, 189–200. https://doi.org/10.1111/j.1467-9922.2012.00712.x

DeKeyser, R., Alfi-Shabtay, I., & Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, *31*, 413–438. https://doi.org/10.1017/S0142716410000056

DeKeyser, R., Alfi-Shabtay, I., Ravid, D., & Shi, M. (2017). The role of salience in the acquisition of Hebrew as a second language: Interaction with age of acquisition. In S. Gass & P. Spinner (Eds.), *Salience in second language acquisition* (pp. 131–146). Abington, UK: Routledge. https://doi.org/10.4324/9781315399027/7.pdf

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In K. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Malden, MA: Blackwell. https://doi.org/10.1002/9780470756492.ch4

Ellis, N. C., & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology A*, *49*, 234–250. https://doi.org/10.1080/027249896392883

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, *54*, 227–275. https://doi.org/10.1111/j.1467-9922.2004.00255.x

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2009). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Edler, R. Erlam, J. Philp, & H. Reinders (Eds.), 2009). *Implicit and explicit knowledge in second language learning, testing, and teaching*. Bristol, UK: Multilingual Matters. https://doi.org/10.21832/9781847691767

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review: DR*, *37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002
Erlam, R. (2006). Elicited imitation as a measure of l2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*, 464–491. https://doi.org/10.1093/applin/aml001
Foster, P., Bolibaugh, C., & Kotula, A. (2013). Knowledge of nativelike selections in a L2. *Studies in Second Language Acquisition*, *36*, 101–132. https://doi.org/10.1017/s0272263113000624
French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, *29*, 463–487. https://doi.org/10.1017/S0142716408080211
Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*, 117–125. https://doi.org/10.1016/j.tics.2014.12.010
García Mayo, M. del P., & García Lecumberri, M. L. (2003). *Age and the acquisition of English as a foreign language*. Clevedon, UK: Multilingual Matters. https://doi.org/10.21832/9781853596407-001
Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, *13*, 65–77. https://doi.org/10.1002/(SICI)1099-0720(199902)13:1/65::AID-ACP548/3.0.CO;2-O
Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*, 269–297. https://doi.org/10.1017/S0272263114000850
Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*, 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9
Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*, 447–474. https://doi.org/10.1016/S0019-9958(67)91165-5
Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*, 665–703. https://doi.org/10.1111/lang.12018
Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*, 311–343. https://doi.org/10.1177/0267658312461497
Grey, S., Williams, J. N., & Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition*, *36*, 611–645. https://doi.org/10.1017/S0272263113000727
Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality.

Learning and Individual Differences, *38*, 44–53. https://doi.org/10.1016/j.lindif.2015.01.019
Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*, 423–449. https://doi.org/10.1017/S0272263113000041
Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1570–1582. https://doi.org/10.1037//0096-1523.26.5.1570
Isbilen, E., Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2018). Bridging artificial and natural language learning: Comparing processing-and reflection-based measures of learning. In C. Kalish, M. Rau, J. Zhu, & T. T. Rogers (Eds.), *Proceedings of the 40th annual conference of the Cognitive Science Society* (pp. 1856–1861). Austin, TX: Cognitive Science Society. Retrieved from https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2622770
Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. J. Davelaar (Eds.), *Proceedings of the 39th annual conference of the Cognitive Science Society* (pp. 564–569). Austin, TX: Cognitive Science Society. Retrieved from https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2503197
Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning*, *67*, 631–664. https://doi.org/10.1111/lang
Jimenez, L. (2003). Intention, attention, and consciousness in probabilistic sequence learning. In L. Jimenez (Ed.), *Attention and implicit learning. Advances in Cognitive Research* (Vol. *48*, pp. 43–68). Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/aicr.48.06jim
Jiménez, L., Mendez, C., & Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 948–969. https://doi.org/10.1037/0278-7393.22.4.948
Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning*, *42*, 217–248. https://doi.org/10.1111/j.1467-1770.1992.tb00708.x
Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/2920538
Karmiloff-Smith, A. (1979). Micro- and macrodevelopmental changes in language acquisition and other representational systems. *Cognitive Science*, *3*, 91–117. Retrieved from https://onlinelibrary.wiley.com/doi/10.1207/s15516709cog0302_1

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science* (Vol. *234*). Cambridge, MA: The MIT Press.

Karpicke, J. D., & Pisoni, D. B. (2004). Using immediate memory span. *Memory & Cognition*, *32*, 956–964. https://doi.org/10.3758/BF03196873

Kaufman, S. B., Deyoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340. https://doi.org/10.1016/j.cognition.2010.05.011

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*, 171–184. https://doi.org/10.1037/a0025405

Kim, J.-E., & Nam, H. (2017). Measures of implicit knowledge revisited: Processing modes, time pressure, and modality. *Studies in Second Language Acquisition*, *39*, 431–457. https://doi.org/10.1017/S0272263115000510

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, *24*, 35–63. https://doi.org/10.1177/0267658307082981

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Edler, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Clevedon, UK: Multilingual Matters. https://doi.org/10.21832/9781847691767-006

Meara, P. M. (2005). LLAMA Language Aptitude Tests *[Measurement instrument and computer software]*. Swansea, UK: Lognostics. Retrieved from http://www.lognostics.co.uk/tools/llama

Marsden, E. J., Williams, J., & Liu, X. (2013). Learning novel morphology: The role of meaning and orientation of attention at initial exposure. *Studies in Second Language Acquisition*, *35*(4), 1–36. https://doi.org/10.1017/S0272263113000296

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*, 395–423. https://doi.org/10.1017/S0142716400003064

Melton, A. W. (1963). Memory. *Science*, *140*, 82–86. https://doi.org/10.1126/science.140.3562.82

Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13310704

Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In D. Luce (Ed.), *Handbook of mathematical psychology* (pp. 2–419). Hoboken, NJ: Wiley. Retrieved from https://philpapers.org/rec/MILFMO

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*, 302–331. https://doi.org/10.1111/j.1467-9922.2010.00626.x

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, *1*(31), 1–9. https://doi.org/10.3389/fpsyg.2010.00031

Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching*, *49*, 113–133. https://doi.org/10.1515/iral.2011.006

Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, *35*, 463–482. https://doi.org/10.1093/applin/amu024

Paradis, M. (2009). *Declarative and procedural determinants of second languages, Studies in bilingualism* (Vol. 40). Amsterdam, The Netherlands: John Benjamins. https://doi.org/10.1075/sibil.40

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology. General*, *119*, 264–275. https://doi.org/10.1037/0096-3445.119.3.264

Pfenninger, S. E., & Singleton, D. (2017). *Beyond age effects in instructional L2 learning: Revisiting the age factor*. Clevedon, UK: Multilingual Matters. https://doi.org/10.21832/9781783097630

Pfenninger, S. E., & Singleton, D. (2019). Starting age overshadowed: The primacy of differential environmental and family support effects on second language attainment in an instructional context: Age and second language attainment. *Language Learning*, *69*, 207–234. https://doi.org/10.1111/lang.12318

Philp, J. (2009). Pathways to proficiency: Learning experiences and attainment in implicit and explicit knowledge of English as a second language. In R. Ellis, S. Loewen, C. Edler, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 167–193). Clevedon, UK: Multilingual Matters. https://doi.org/10.21832/9781847691767-010

Pierrehumbert, J. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics* (pp. 177–228). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/5582.001.0001

Plonsky, L., Marsden, E., Crowther, D., Gass, S., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, *36*, 583–621. https://doi.org/10.1177/0267658319828413

Core Team, R. (2018). *R: A language and environment for statistical computing [Computer software]*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology. General*, *118*, 219–236. https://doi.org/10.1037/0096-3445.118.3.219

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*, 829–856. https://doi.org/10.1017/S0142716411000580

Rebuschat, P., & Williams, J. N. (2013). Implicit learning in second language acquisition. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Oxford, UK: Blackwell. Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0529

Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, *72*, 67–109. https://doi.org/10.1016/s0010-0277(99)00031-1

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/10.1016/s0010-0277(98)00075-4

Service, E. (1992). Phonology, working memory, and foreign-language learning. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *45*(1), 21–50. https://doi.org/10.1080/14640749208401314

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711), 20160059. https://doi.org/10.1098/rstb.2016.0059

Shiu, L.-J., Yalçın, Ş., & Spada, N. (2018). Exploring second language learners' grammaticality judgment performance in relation to task design features. *System*, *72*, 215–225. https://doi.org/10.1016/j.system.2017.12.004

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford, UK: Oxford University Press.

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*, 723–751. https://doi.org/10.1111/lang.12129

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*, 747–790. https://doi.org/10.1111/lang.12241

Tarone, E. (1988). *Variation in interlanguage*. London, UK: Hodder Arnold.

TextAloud (Version 4) [Computer software]. (2009). Clemmons, NC: NextUp Technologies.

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the*

Royal Society of London. Series B, Biological Sciences, *372*(1711), 1–10. https://doi.org/10.1098/rstb.2016.0056

Thompson-Schill, S. L., Ramscar, M., & Chrysikou, E. G. (2009). Cognition without control: When a little frontal lobe goes a long way. *Current Directions in Psychological Science*, *18*, 259–263. https://doi.org/10.1111/j.1467-8721.2009.01648.x

Vafaee, P., Suzuki, Y., & Kachisnke, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, *39*, 59–95. https://doi.org/10.1017/S0272263115000455

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology*, *141*, 65–82. https://doi.org/10.1016/j.jecp.2015.06.015

Verhagen, J., Leseman, P., & Messer, M. (2015). Phonological memory and the acquisition of grammar in child L2 learners. *Language Learning*, *65*, 417–448. https://doi.org/10.1111/lang.12101

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition*, *21*, 1–48. https://doi.org/10.1017/S0272263199001011

Williams, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–353). Bingley: Castle Hill, NSW, Australia: Emerald Press.

Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, *118*, 522–553. https://doi.org/10.1016/j.lingua.2007.03.003

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*, 457–486. https://doi.org/10.1017/S0272263114000370

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Exact Probability Estimates for Pearson Correlations.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

The Roles of Age, Memory, and Implicit Learning in Immersed and Instructed Grammar Learning

What This Research Was About and Why It Is Important

People who immigrate as children tend to reach higher levels of second language proficiency than those who immigrate as adults. Some studies have suggested that this is because children mainly rely on a type of unconscious learning (implicit learning) that adults no longer have access to. This study tested this idea by measuring the grammar knowledge of long-term second language learners who had learned their language either as immigrants or in a classroom, and who had started learning at a wide range of ages. We also took cognitive measures of short-term memory and implicit learning. We found that only immersed learners' grammar knowledge showed signs of having been learned implicitly, and this was regardless of the age they had started learning.

What the Researchers Did

- Participants were 71 native speakers of Polish who learned English as a second language. Half of them (35) learned English when they immigrated to the United Kingdom, and half (36) learned as foreign language students in Poland. We also tested 30 native speakers of British English.
- All Polish participants were fluent daily users of both English and Polish, with English at an upper-intermediate or advanced level. They had a minimum of 12 years of exposure to English, and had started learning English between the ages of 1 year and 35 years.
- We tested all participants' grammar knowledge by asking them to listen to 110 English sentences and tell us whether they were correct. Half of the sentences contained errors. Polish participants were also given two cognitive tests: a short-term memory test, and an implicit-learning test.
- We compared how the two groups performed on the grammar test, and how it was related to the age they started learning English, and to their short-term memory and implicit learning.

What the Researchers Found

- For Polish immigrants to the U.K., people who had immigrated as children had better grammar knowledge than those who came as teenagers, who were better than those who came as adults. For instructed learners of English in Poland, age was not as important.

- For Polish immigrants to the U.K., short-term memory and implicit learning abilities were related to how well they performed on the grammar test. For instructed learners in Poland, memory and implicit learning were not related to their performance on the grammar test.
- For Polish immigrants to the U.K., having a good short-term memory was more important for those who had started when they were teenagers or adults. Implicit learning ability was important regardless of the age they started.

Things to Consider

- The role of the age at which one begins to learn a language, short-term memory, and implicit learning was different for immigrant learners compared to classroom learners. This suggests that learning in immersion situations is different from classroom learning, and we should be cautious in applying findings from immigrant studies to the classroom.
- In immigrant learners, the importance of implicit learning ability was the same for all learners, regardless of the age they started. This suggests that children and adults can use the same type of learning if they are immersed in the language.
- Finally, in immigrant learners, having a good short-term memory was more important for those who started as teens or adults. It seems that having a good short-term memory can to some extent help "make up for" starting learning later.

Materials, data, open access article: Materials and data are publicly available at: https://iris-database.org and https://doi.org/10.17605/OSF.IO/GCMXK

How to cite this summary: Bolibaugh, C., & Foster, P. (2021). The roles of age, memory, and implicit learning in immersed and instructed grammar learning. *OASIS Summary* of Bolibaugh & Foster (2021) in *Language Learning*. https://oasis-database.org